

Forecasting the Transmission Trends of Respiratory Infectious Diseases with an Exposure-Risk-Based Model at the Microscopic Level

Ziwei Cui, Ming Cai, Yao Xiao, Zheng Zhu, and Mofeng Yang

Abstract—Respiratory infectious diseases (e.g., COVID-19) have brought huge damages to human society, and the accurate prediction of their transmission trends is essential for both the health system and policymakers. Most related studies concentrate on epidemic trend forecasting at the macroscopic level, which ignores the microscopic social interactions among individuals. Meanwhile, current microscopic models are still not able to sufficiently decipher the individual-based spreading process and lack valid quantitative tests. To tackle these problems, we propose an exposure-risk-based model at the microscopic level, including 4 modules: individual movement, virion-laden droplet movement, individual exposure risk estimation, and prediction of new cases. First, the front two modules reproduce the movements of individuals and the droplets of infectors' expiratory activities. Then, the outputs are fed to the third module for estimating the personal exposure risk. Accordingly, the number of new cases is predicted in the final module. Our model outperforms 4 existing macroscopic or microscopic models through the forecast of new cases of COVID-19 in the United States. Specifically, mean absolute error, root mean square error and mean absolute percentage error by our model are 2454.70, 3170.51, and 3.38% smaller than the minimum results of comparison models, respectively. In sum, the proposed model successfully describes the scenarios from a microscopic perspective and shows great potential for predicting the transmission trends with different scenarios and management policies.

Index Terms—Respiratory Infectious Diseases, Microscopic Model, Exposure Risk, COVID-19

I. INTRODUCTION

THE unexpected outbreak and rapid spread of Corona Virus Disease 2019 (COVID-19) have damaged the world

This work was supported by the National Natural Science Foundation of China (Grant No. 72101276), and the Fundamental Research Funds for the Central Universities (Grant No. 2242021R10037)

Ziwei Cui, Ming Cai, and Yao Xiao are with the School of Intelligent System Engineering, Sun Yat-Sen University, Shenzhen, Guangdong, China (e-mail: cuizw3@mail2.sysu.edu.cn; caiming@mail.sysu.edu.cn; xiaoyao9@mail.sysu.edu.cn).

Zheng Zhu is with the College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, Zhejiang, China (e-mail: zhuzheng89@zju.edu.cn).

Mofeng Yang is with Maryland Transportation Institute, Department of Civil and Environmental Engineering, University of Maryland at College Park, Maryland, USA (e-mail: mofeng@umd.edu).

Corresponding author: Yao Xiao (e-mail: xiaoyao9@mail.sysu.edu.cn)

tremendously and brought a profound influence on people's lives [1]–[3]. At present, a growing number of countries have moved into a post-pandemic phase, i.e., the overall spread of COVID-19 has been controlled, but there have been intermittent small-scale outbreaks [4]. COVID-19 is a respiratory infectious disease (RID) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [5]. RID poses a severe threat to the population and public health, and recent studies have found that the loss of life expectancy due to RID stood at 1.29 years in 2017 globally [6]. To live with the ongoing challenges presented by COVID-19 or the re-emergence of other respiratory viral infections, people must take the awareness and measures of infection prevention and control into daily life [7].

Vaccines and non-pharmaceutical interventions (NPIs) play essential roles in stopping the spread of respiratory viral infections. For a new RID, vaccine development needs a long cycle [8]. The vaccines' effectiveness would also be reduced by the side effects of serial vaccination [9]–[11]. Moreover, the virus of the RID can mutate quickly into many variations, such as VUI-202012/01 that is just one of many variations of the SARS-CoV-2, and whether the vaccines are effective against these new variants is unknown [12]. On the other hand, NPIs (e.g., maintaining safe social distance, face masks [13], [14]) efficiently respond to emerging epidemics, and they are long-term approaches to regulate people's social behavior [15], [16]. In a public place, it is important to find the most critical NPIs because they bring the lowest number of infected cases. Thus, a model that can accurately describe diseases' transmission process is necessary to help evaluate NPIs by estimating the corresponding disease transmission risks and trends [17].

Mathematical epidemiological models are helpful to decipher the complex transmission process of epidemics [18], which generally includes macroscopic level models and microscopic level models [19]. At the macroscopic level, researchers focus on the infection and recovery process of disease among the population [20], and have proposed many compartmental models and computational intelligence methods: in the former, there are susceptible–infected–removed, or recovered (SIR) model [21], the susceptible–infected–recovered–susceptible (SIRS) model [22], and their extended modifications [23]–[25]; in the latter, there are the hybrid intelligent approach

based on fractal theory and fuzzy logic [26], the multiple ensemble neural network model with fuzzy response aggregation [27], the hybrid intelligent fuzzy fractal approach [28], and others. Although the macroscopic models have established the research discipline of mathematical epidemiology, most of them need continuously updated data or an enormous amount of data to obtain optimized system parameters, and they may ignore details in modeling the social interactions among individuals. Therefore, macroscopic approaches can be insensitive in evaluating NPIs or require strong assumptions to overcome the incapability. Fortunately, microscopic models partially address these limitations because they focus on the disease spreading between individuals. For example, scholars have estimated individual-level spatiotemporal travels between several locations based on daily activities (such as work, study, or shopping) [29]. In these models, RIDs can transmit between individuals in the exact location simultaneously and spread between locations due to individuals' movements. It should be noted that the location mentioned here is a particular area like a school or shopping mall, not the exact physical location of the individual. Unlike these activity-based models, some studies focus on respiratory viruses' transmission and construct the Well-Riley models and aerosol infectious dose-response models [30], [31]. These models assume that infectious particles are distributed homogeneously, which results in the same infection risks among individuals regardless of their physical distance from the infector. Recent studies have considered the spatiotemporal distribution of pathogens in the environment under different transmission routes, and the personal exposure risk is determined based on the duration and distance of infectious contacts [32], [33]. However, transmission is modeled as a static event between infectors and susceptible individuals. Thus, these models cannot describe the epidemic spreading process with time-varying individual physical distances during individual movements. Nevertheless, pedestrian dynamics is suitable for describing individual decisions and actions in mass gathering scenarios. Previous studies have integrated these decisions and actions into epidemic spreading models to counter the aforementioned problems [34], [35].

Several studies have explored pedestrian-based epidemic spreading models, which are composed of the individual movement module (e.g., the social force model [19], the nomad model [15], a social force model coupled with an Eikonal equation [36]) and the disease transmission module (e.g., the model based on the cut-off distance [19], the QVEmod [15], the non-local SEIS contagion model [36]). The former module simulates the general individual crowd movement and outputs time series of personal positions. Based on the outputs from the first module, the second module evaluates the disease transmission risk from the infectors to susceptible individuals. However, these models ignored critical factors for simplification, e.g., the exposure risk is an invariable value when the individual is exposed to the infector's influence area [19], and they potentially overestimate or underestimate the number of high-risk exposed people. Besides, only specific situations are analyzed, such as the cruise [37], the supermarket [38], and the academic building [39]. Moreover, different scenarios have various scales and geometries, e.g., a restaurant whose

dimensions are 10 m \times 9 m has tables, chairs, the cash register, and other objects in the case of [15]. These changeable settings affect the virus transmission risks and bring more model inputs and computation costs. On the other hand, a general public place can represent all scenes, e.g., an empty room of 10 m \times 9 m represents a restaurant or store of the same scale without considering the indoor geometries. Meanwhile, most outputs of existing microscopic-level models are results without actual values to verify, such as the personal exposure risk and the number of high-risk exposed people. Then, these models are used directly or after the qualitative analysis but lack quantitative tests [19], [40]. Hence, building a model for a general public place that can be validated and applied to all scenes is necessary and practical.

In this study, an exposure-risk-based model at the microscopic level is developed, and the principal contributions can be described as follow. *First*, the movements of individuals as well as the dynamics of droplets are respectively modeled and coupled to calculate the personal exposure risk, hence an integrated transmission process of the RID at the individual level is formulated. *Second*, our model can predict the spreading trends from a microscopic perspective of individuals, which indicates that many specific factors like scene (e.g., geometry), activity (e.g., point of interest) and strategy (e.g., travel restriction) in the real-world scenarios can be captured and further investigated. *Third*, a bridge between the macroscopic epidemic transmission data and our microscopic model at individual level is built. As a result, our model can be quantitatively calibrated and validated through the macroscopic data, such as the number of new cases.

The rest of the paper is organized as follows. Section II introduces our model, and Section III manifests applications of the model based on real-world data in the COVID-19 of the United States. Section IV discusses the comparison models and ours. Finally, we summarize this paper and present suggested directions for future work in Section V.

II. MODEL

To construct the microscopic level exposure-risk-based model, we should define the exposure risk of individuals. RIDs are transmitted through viral droplets produced by respiratory activities such as breathing, talking, coughing, and sneezing. Here, the typical symptom of most RIDs, i.e., coughing, is adopted in our model. Thus, we define the instantaneous personal exposure risk as the maximal mass of virion-laden droplets produced by a typical cough exposed to infectors at a certain moment. Then, the personal exposure risk can be calculated by directly summing up instantaneous personal exposure risks during the dwell time.

The components of our model are shown in Fig. 1. There are three important parameters (i.e., number of individuals C^{total} , number of infectors among individuals C^{inf} , and mean dwell time of individuals T^{dwell}) as the input. Besides, the output includes not only the result of existing pedestrian-based spreading models (i.e., number of high-risk exposed people C^{risk}), but also the key outcome (i.e., number of new cases C^{new}). In addition, personal exposure risk is obtained as the intermediate output.

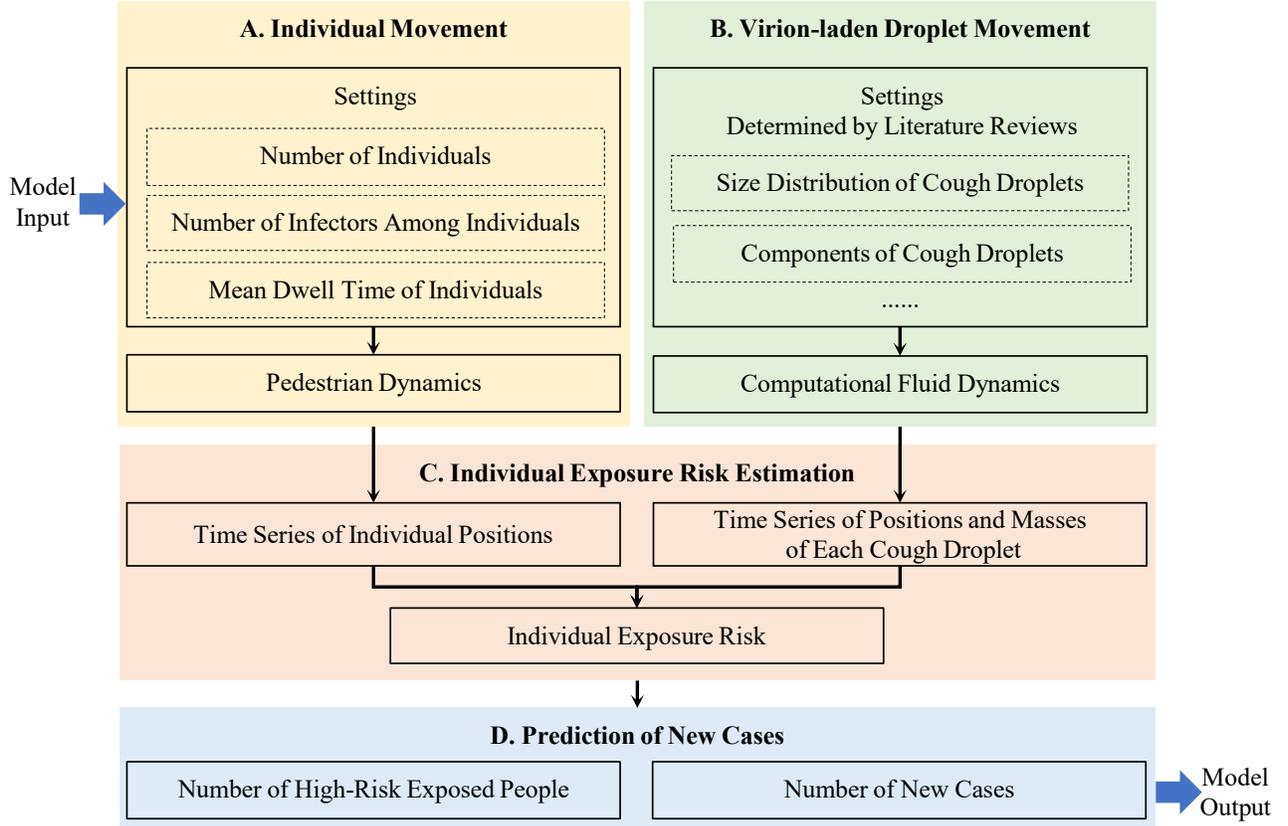


Fig. 1. The flowchart of our model.

A. Individual Movement

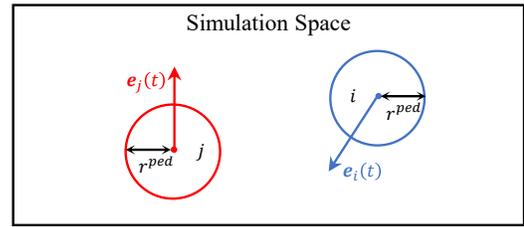
In this part, the movements of individuals are modeled within a general public place. *First*, we simulate the space and individuals of the reality. *Second*, the pedestrian dynamic model (i.e., the social force model) to reproduce individuals' movements is introduced. *Finally*, time series of individual positions can be estimated for Section II-C.

Herein, the simulation space is a general place without obstacles, but it has boundaries like an empty indoor room. Whether individuals from model inputs are infected or not, their movement modelings are the same, as shown in Fig. 2, where each person is represented as a circle with a radius r^{ped} . Thus, when we model the motion of one person during the dwell time, the movements of all others can be determined in the same way.

The social force model is one of the most widely used microscopic models, and it is also the basic model to simulate pedestrian dynamics in commercial software (e.g., PTV-Vissim, MassMotion) [41], [42]. Hence, the social force model is applied here, and the movement of individual i (with a mass of m_i) at time t is driven by the resultant force $\mathbf{f}_i(t)$ as

$$\mathbf{f}_i(t) = \mathbf{f}_i^{\text{drv}}(t) + \sum_{j^{\text{near}}} \mathbf{f}_{i,j^{\text{near}}}^{\text{ped}}(t) + \sum_w \mathbf{f}_{i,w}^{\text{obs}}(t), \quad (1)$$

where $\mathbf{f}_i^{\text{drv}}(t)$ reflects the desire of individual i to maintain a certain walking speed v_i^0 towards a certain direction in a relaxation time τ_i ; $\mathbf{f}_{i,j^{\text{near}}}^{\text{ped}}(t)$ is the interaction force between



• Infectors • Susceptible Individuals — Boundaries

Fig. 2. The sketch map of the individuals moving in the simulation space.

the objective individual i and the neighboring individual j^{near} ; $\mathbf{f}_{i,w}^{\text{obs}}(t)$ means the instant interaction force between the objective individual i and the wall/obstacle w . More details about the social force model can be found in the literature [43]–[45].

As the motion of individual i is driven under Newton's second law with a second-order dynamics function, the velocity $v_i(t)$ and the unit direction vector $\mathbf{e}_i(t)$ of individual i is determined in

$$\frac{dv_i(t)\mathbf{e}_i(t)}{dt} = \frac{\mathbf{f}_i(t)}{m_i}, \quad (2)$$

based on which the location $\mathbf{p}_i(t)$ of individual i is finally

obtained as

$$\frac{d\mathbf{p}_i(t)}{dt} = v_i(t)\mathbf{e}_i(t). \quad (3)$$

In addition, to simulate the crowd movements, a random walking process is set by adjusting the desired initial direction \mathbf{e}_i^0 . Then, when individual i is close to hitting the boundary, the model will force the person to change the desired direction to any direction \mathbf{e}_i^b that is away from the boundary.

B. Virion-Laden Droplet Movement

Virion-laden droplet transmission of a typical cough is modeled through Computational Fluid Dynamics (CFD) simulations within a closed environment. *First*, the computational domain and grids are simulated. *Second*, we choose the appropriate methods and parameter settings from literatures for the numerical simulation. *Finally*, after completing the simulation with the commercial CFD solver ANSYS Fluent release 2020 R2, time series of positions and masses of each cough droplet can be determined for Section II-C.

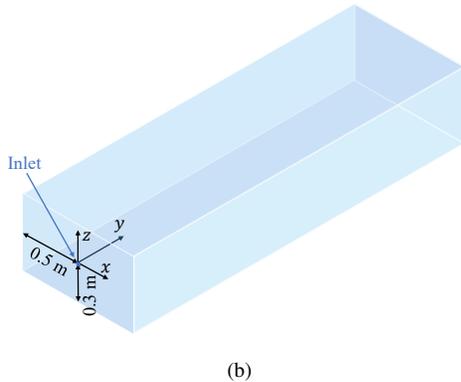
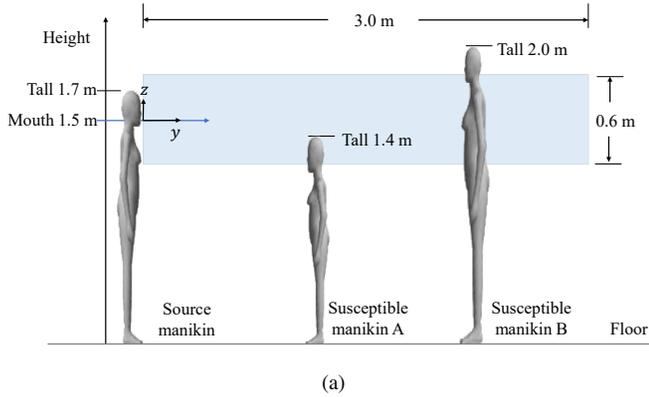


Fig. 3. Schematic diagram of (a) the computational domain with source and susceptible manikins, and (b) the numerical simulation computational domain.

An infector is represented by a source manikin 1.70 m tall, representing an average-sized human as Fig. 3(a) [46]. The mouth of the infector is 1.50 m from the ground, and the mouth outflow is roughly horizontal. For susceptible people whose height ranges from 1.40 m to 2.00 m, we take 1.20 m to 1.80 m as the breathing height area, as shown in the blue area of Fig. 3(a). The computational domain is in a cuboid

shape and is illustrated by Fig. 3(b), in which “x”, “y”, and “z” axes represent the lateral, axial (streamwise), and vertical directions, respectively. The length of the computational domain in the “y” axis is set to be 3.00 m to consider the scope of a cough fully. Now, the blue area of Fig. 3(a) is the view of Fig. 3(b) along the “x” axis. The inlet is a square with an area of $3.70 \times 10^{-4} \text{ m}^2$, representing the human mouth [47]. The center of the inlet is denoted by (0, 0, 0), which is also the origin of the coordinate system. The computational domain and grids are generated by using Gambit 2.4.6. The grid resolution is 0.01 m, and the total number of computational cells is approximately 1.80 million.

To accurately estimate the consequences of the coughing event, reliable modeling methods and settings of numerical simulation are important. The transmission medium of the cough is modeled as an incompressible ideal gas with constant properties calculated at ambient conditions. The flow field evolution of coughing is time-dependent, so the simulations are conducted under a transient condition. The gravitational acceleration is -9.81 m/s^2 along the “z” axis, and the energy equation is required. Since the droplet volume fraction is very low in the cough flow, the Eulerian-Lagrangian method is used in this study [48], [49]. Similar to the study [47], a time step size of 0.01 s is used, with 10 sub-iterations. The total flow time of the transient simulation is 15.00 s, which is enough to investigate the dynamic characteristics of the droplets produced by coughing. More detailed settings are determined based on literature [50]–[56].

The droplet dispersion process in the computational domain is obtained after the simulation. The results show that the drop velocity increases with a larger diameter, and the droplet can be suspended for a longer period when the diameter is less than or equal to $1.00 \times 10^{-5} \text{ m}$, as the same in [47].

C. Individual Exposure Risk Estimation

In this part, we *first* give the mathematical presentation of the instantaneous personal exposure risk. As mentioned above, when the person and the cough droplets meet in the same simulation place, the instantaneous exposure risk is defined as the possible maximal mass of droplets suffered. *Then*, based on the numerical simulation results of the typical cough, we count the mass of droplets vary with the time at several fixed distances planes, and analyze the data distribution to determine the model function, which is the final formula of the instantaneous exposure risk. *Finally*, the total exposure risk during the visiting time for each person is estimated.

It should be noted that, according to a few studies that analyze the dispersion of cough-generated droplets in the wake of a walking person [57], the situation can be more complicated when there are more individuals. Besides, the direction of the cough is likely to be different from that of the individual’s movement, so the impact of walking on cough transmission is unpredictable. Therefore, in this model, we assume that the infector j is stationary when he/she starts the g -th cough in the position $\mathbf{p}_{j,g}^{\text{start}}$ at time $t_{j,g}^{\text{start}}$. Then, cough droplets’ movements are only determined by the physical distance from $\mathbf{p}_{j,g}^{\text{start}}$ and the time interval from $t_{j,g}^{\text{start}}$.

All positions needed here can be determined based on the crowd simulation in Section II-A. When individual i in the position $\mathbf{p}_i(t)$ at time $t(\geq t_{j,g}^{\text{start}})$, the instant distance from the position $\mathbf{p}_{j,g}^{\text{start}}$ is $d_{i,j,g}(t)$ and the time interval from $t_{j,g}^{\text{start}}$ is $t_{j,g}^{\text{interval}} = t - t_{j,g}^{\text{start}}$. If $d_{i,j,g}(t)$ and $t_{j,g}^{\text{interval}}$ are both in the cough infectious distance D^{inf} and infectious time T^{inf} respectively, based on the definition, we have

$$E_{i,j,g}(t) = M_{j,g}(d_{i,j,g}(t), t_{j,g}^{\text{interval}}), \quad (4)$$

where $E_{i,j,g}(t)$ is the instantaneous exposure risk of the individual i exposed to individual j 's g -th cough, and $M_{j,g}(d_{i,j,g}(t), t_{j,g}^{\text{interval}})$ is the maximal mass of droplets generated from individual j 's g -th cough when spreading to the distance $d_{i,j,g}(t)$ after $t_{j,g}^{\text{interval}}$.

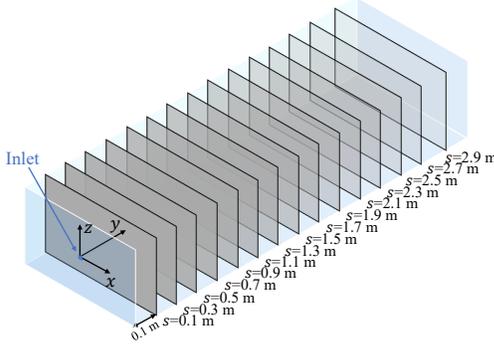


Fig. 4. 15 representative x-z planes in the three-dimensional schematic diagram of the computational domain.

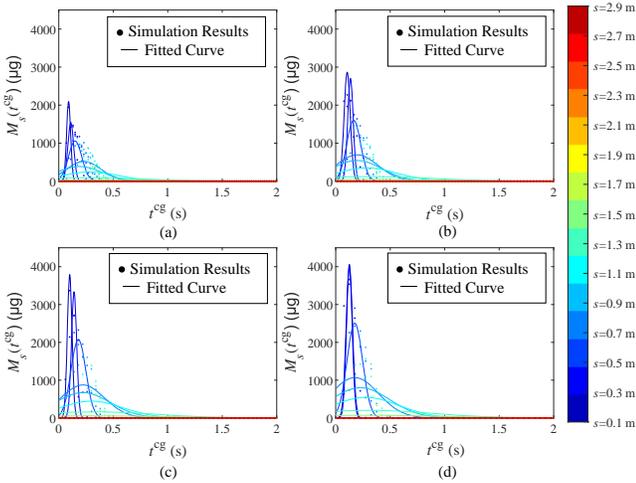


Fig. 5. $M_s(t^{\text{cg}})$ versus t^{cg} at different times: (a) $T^{\text{step}} = 0.02\text{s}$; (b) $T^{\text{step}} = 0.03\text{s}$; (c) $T^{\text{step}} = 0.04\text{s}$; (d) $T^{\text{step}} = 0.05\text{s}$.

On the assumption that all coughs are typical, to get the deterministic mathematical formula of $M_{j,g}(d_{i,j,g}(t), t_{j,g}^{\text{interval}})$, it is sufficient to analyze the change of droplet mass during the transmission of one typical cough. Since the density of all droplets in Section II-B is set to the same value, the larger the diameter, the greater the droplet mass. Droplets with a large mass fall quickly, and their destinations are always close to the injector. In contrast, others with small mass exist in

the computation region for a long time, and their physical distances from the injector are random. Hence, it is hard to find a suitable deterministic formula to describe the change of droplet mass with the time-varying distance. To solve this problem, we select representative x-z planes and model the arriving droplet mass varies with time for each plane. For a typical cough, the mass of droplets $M_s(t^{\text{cg}})$ passing through the x-z plane along “y” axis in distance s m at time $t^{\text{cg}}(\leq T^{\text{inf}})$ can be counted. Since the individual is represented as a circle of 0.2 meters (i.e., $r^{\text{ped}} = 0.1$ m) [19], the calculation domain in the simulation is divided with a length of 0.20 m along the “y” axis, and the x-z plane at the center of each segment is taken as the representative of each region. More exactly, the value of s starts from 0.10 m to 2.90 m with an interval of 0.20 m, and there are 15 representative x-z planes $\Phi = \{\phi_1, \phi_2, \dots, \phi_s, \dots, \phi_{15}\}$ as shown in Fig. 4. Moreover, to avoid reducing the counted droplet mass caused by the setting of timestep T^{step} in pedestrian dynamics simulation, the module uses the total mass of droplets passing through ϕ_s during $\lceil \frac{t^{\text{cg}}}{T^{\text{step}}} \rceil$ timesteps as $M_s(t^{\text{cg}})$. As settings in the second module (Section II-B), the simulation time step size of cough droplet transmission is 0.01 s. Therefore, the simulation time step size of pedestrian dynamics T^{step} should be bigger than 0.01 s. When we set $T^{\text{step}}=0.02$ s, 0.03 s, 0.04 s, 0.05 s, respectively, the statistical results show that the distribution of $M_s(t^{\text{cg}})$ varies with t^{cg} and it conforms to the Gaussian distribution (see Fig. 5). Thus, for the ϕ_s , $M_s(t^{\text{cg}})$ is obtained by

$$M_s(t^{\text{cg}}) = a_s * \exp\left(-\left(\frac{t^{\text{cg}} - b_s}{c_s}\right)^2\right), \quad (5)$$

where a_s , b_s and c_s are parameters of the fitted Gaussian distribution function for specific ϕ_s .

After determining the function of the $M_s(t^{\text{cg}})$ for a typical cough, since t^{cg} in the typical cough is the same as $t_{j,g}^{\text{interval}}$ of individual j 's g -th cough, if $d_{i,j,g}(t) \in [s - 0.1, s + 0.1)$ and $t_{j,g}^{\text{interval}} \leq T^{\text{inf}}$, the formular of $M_{j,g}(d_{i,j,g}(t), t_{j,g}^{\text{interval}})$ is represented by

$$M_{j,g}(d_{i,j,g}(t), t_{j,g}^{\text{interval}}) = a_s * \exp\left(-\left(\frac{t_{j,g}^{\text{interval}} - b_s}{c_s}\right)^2\right). \quad (6)$$

Consequently, if $d_{i,j,g}(t) \in [s - 0.1, s + 0.1)$ and $t \leq t_{j,g}^{\text{start}} + T^{\text{inf}}$, there is

$$E_{i,j,g}(t) = a_s * \exp\left(-\left(\frac{t - t_{j,g}^{\text{start}} - b_s}{c_s}\right)^2\right). \quad (7)$$

Then, the exposure risk of individual i during the visit follows

$$E_i = \sum_{t=t_i^{\text{enter}}}^{t_i^{\text{enter}}+T_i^{\text{dwell}}} \sum_{j=1}^{J(t)} \sum_{g=1}^{J_G(t)} E_{i,j,g}(t), \quad (8)$$

where t_i^{enter} and T_i^{dwell} respectively denotes the place enter time and the dwell time of individual i , $J(t)(\leq C^{\text{inf}})$ indicates the number of infectors in the simulation place at time t , and $J_G(t)$ is the number of infectious coughs of the infector j at time t .

D. Prediction of New Cases

In this part, we first determine the number of high-risk exposed people C^{risk} during the simulation horizon. Then, based on C^{risk} , we present a novel method to predict the number of new cases C^{new} .

Individual i whose $E_i > \alpha$ is defined as the high-risk exposed individual, where α is the cut-line of high exposure risk. When the number of total individuals for the simulation is C^{total} , the value of C^{risk} changes with the α and we have

$$C^{\text{risk}}(\alpha) = \sum_{i=1}^{C^{\text{total}}} \psi(E_i, \alpha), \text{ where } \psi(E_i, \alpha) = \begin{cases} 1, & \text{if } E_i > \alpha \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Admittedly, C^{new} always increases with the growth of C^{risk} , i.e., there is a positive correlation between the C^{new} and the C^{risk} . From the viewpoint of mathematics, C^{new} can be set as a function F of C^{risk} quantitatively as

$$C^{\text{new}} = F(C^{\text{risk}}). \quad (10)$$

We consider the simplest relationship to determine the function $F(C^{\text{risk}})$ by assuming that $F(C^{\text{risk}})$ is a linear equation. Moreover, in an extreme case, when there are no infectious diseases, there are no high-risk exposed people and new cases, i.e., $C^{\text{new}} = C^{\text{risk}} = 0$. Based on these analyses, the function $F(C^{\text{risk}})$ passes through the origin and be defined as

$$C^{\text{new}} = F(C^{\text{risk}}) = \beta * C^{\text{risk}}(\alpha), \quad (11)$$

where β is the proportionality coefficient.

To determine the values of parameters α and β in different scenes, the actual historical numbers of daily new cases are needed. When C_v^{new} and C_v^{real} are the predicted and actual values of the $v(= 1, 2, \dots, V)$ -th historical observation, respectively, indexes such as mean absolute error (MAE) can be used to describe the error between C_v^{new} and C_v^{real} . Herein, we select MAE as the index, and the appropriate values of α and β can be estimated as

$$\alpha^*, \beta^* = \arg \min_{\alpha, \beta} \left(\frac{1}{V} \sum_{v=1}^V |C_v^{\text{new}} - C_v^{\text{real}}| \right). \quad (12)$$

Finally, the C^{new} in the future can be predicted with the α^* and β^* via (9-11).

III. APPLICATIONS IN THE COVID-19 OF THE UNITED STATES

A. Data Sources

To estimate RID transmission based on the actual situation, we select data from all 50 states and Washington, D.C. in the United States (U.S.) during the spreading of COVID-19. From May 1st to July 25th in 2020, 79 days with complete national-level data are utilized, and statistics of public data are made based on days.

There are three inputs in our model, as shown in Fig. 1, and they are introduced as follows: 1) The number of individuals C^{total} each day is approximated as the daily population traveling out of home, which is collected from the U.S. Department of Transportation's Bureau of Transportation Statistics Trips by Distance – National data product [58], [59]. 2) The number

of infectors among individuals C^{inf} needs to be determined, but it cannot be found based on the public data because whether each infector has a trip in a single day is unknown. In fact, there are some undiagnosed cases in the travel crowd, so we present a reasonable assumption, i.e., the ratio of infectors to individuals is equal to the ratio of the total number of cases to the population traveling out of home. The total number of cases in each day is found from the U.S. Centers for Disease Control and Prevention [60]. 3) Mean dwell time T^{dwell} of individuals represents the average time spent for each person in the public space. We collect this data from the University of Maryland COVID-19 Impact Analysis Platform [19], [59].

The study period (79 days selected from May 1st to July 25th) is divided into a training set (the first 60 days, i.e., selected from May 1st to July 6th) and a testing set (the rest 19 days, i.e., from July 7th to July 25th), as shown in Fig. 6. Besides, to reduce the computational cost, the number of individuals C^{total} and the number of infectors among individuals C^{inf} are scaled down with a proportion ρ for the simulation in both the training and testing sets. Consequently, the main outputs of the simulation, i.e., the number of high-risk exposed people C^{risk} and the number of daily new cases C^{new} , will be expanded with the same proportion ρ after the simulation is finished. We use the first day in the study period (May 1st, 2020) as the benchmark, then 245469060 individuals who have at least one trip at that day can be scaled down to 10000 individuals for the simulation, thus the proportion $\rho = \frac{10000}{245469060} = 4.07 \times 10^{-5}$. The values of α^* and β^* can be estimated in the training set, and then be adopted in the testing set to evaluate the model.

B. Simulation Setups

Considering the fact that realistic indoor scenes vary in different places (e.g., restaurant, cinema and subway hub), and the specific geometry or activity might impact the transmission process. Besides, the geometries and activities can be described by our microscopic model. Even so, paying too much attention on the heterogenous settings is not necessary, since the consideration of all specific settings is not practicable, and a more practical way is to formulate a general public space. Accordingly, a 22 m \times 22 m indoor room without obstacles is formulated, and there is an entrance on one side of the room and an exit on the opposite side (see Fig. 7). The red and blue dots represent the infectors and the susceptible individuals, respectively.

In the simulations, the specific individual number, the infector number and the mean dwell time are determined according the realistic data introduced in Section III-A. Initially, no individual exists in the room, and the crowd are designed to enter the indoor room through the entrance in sequence. The individual follows a random walking pattern in the room during the dwell time and leave the room through the exit. The individual is represented by a circle of 0.2 meters, and the relaxation time and the desired speed are $\tau_i = 0.50$ s and $v_i^0 = 1.34$ m/s, respectively. The infector averagely coughs every 15 s after entering the room, and the infectious time of a cough follows a uniform distribution from 0 to 15 s. The

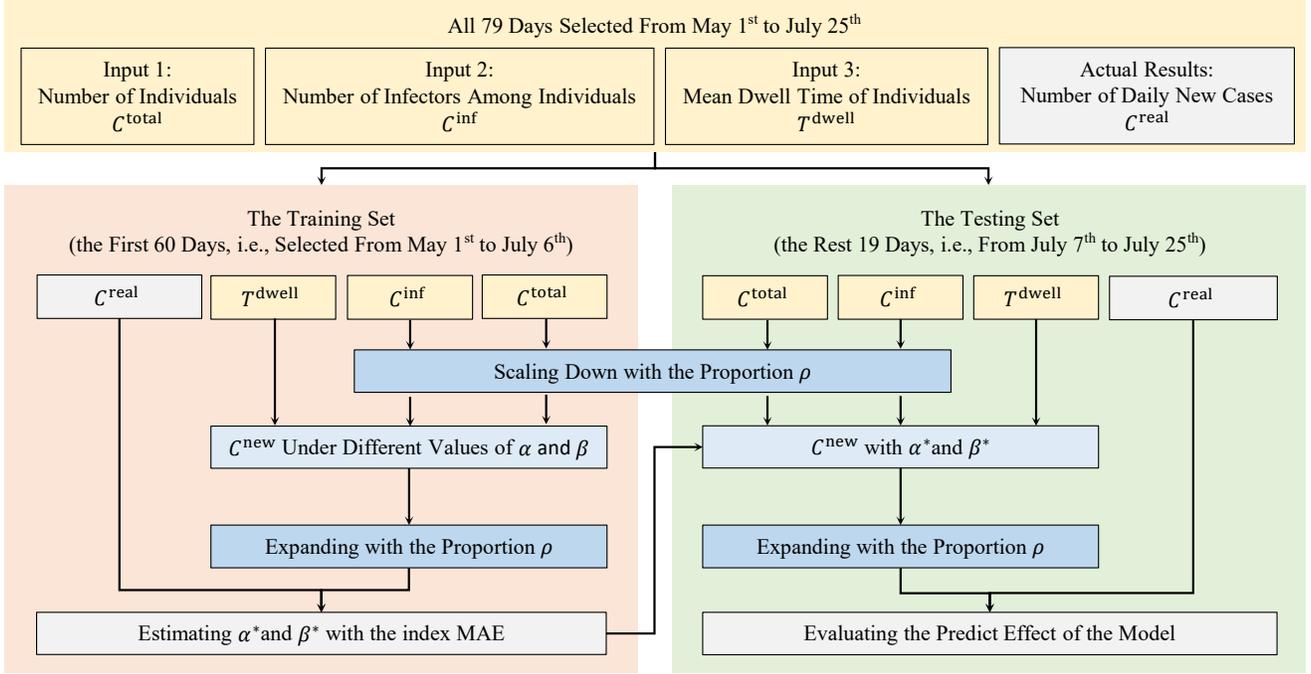


Fig. 6. Schematic diagram of data relations used in this case.

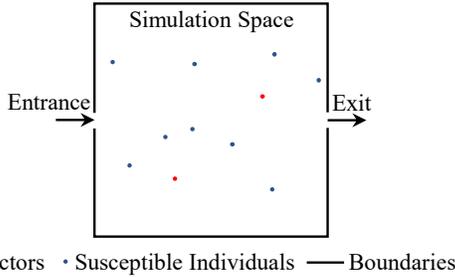


Fig. 7. The sketch map of the simulation space in this case.

TABLE I
EXAMPLES OF SIMULATION INPUTS AND STATISTICAL RESULTS

Date	Simulation Inputs			Statistical Results	
	C^{total}	C^{inf}	T^{dwell} /minutes	C^{sus} with $E_i=0$	C^{sus} with $E_i>0$
May 1 st	10,000	39	26.29	1,092	8,869
May 17 th	10,210	55	24.56	724	9,431
July 10 th	10,210	128	25.64	225	9,857
July 25 th	9,844	169	23.86	138	9,537

infectious distance D^{inf} of a typical cough is set at 1.70 m since $M_s(t^{\text{c}^{\text{g}}})$ equals almost 0 at a greater distance according to Fig. 5(c).

C. Simulation Outputs

1) *Individual Exposure Risk Estimation*: The exposure risk of each person in a single day is obtained from the simulation.

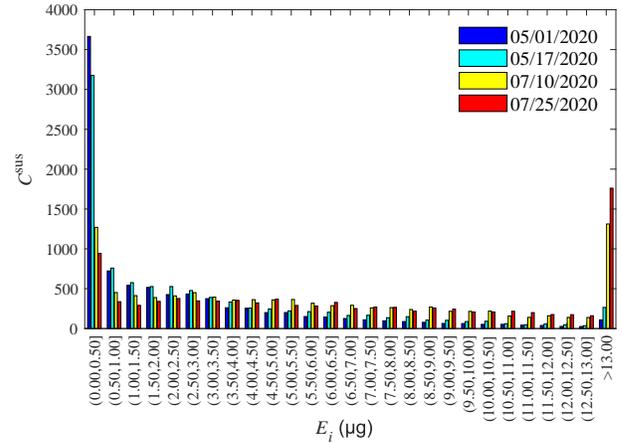


Fig. 8. The number of individuals varies in different exposure risk segments.

Picking 4 days (May 1st, May 17th, July 10th, July 25th) as examples, the inputs of the simulation (i.e., C^{total} , C^{inf} , and T^{dwell}) and the statistical results of simulation outputs (i.e., the number of susceptible individuals C^{sus} with $E_i = 0$ or $E_i > 0$) are shown in Table I and Fig. 8.

According to Table I and Fig. 8, simulation results of our model show that C^{sus} with $E_i \in (0.00, 0.50] \mu\text{g}$ on May 1st is 2722 more than that on July 25th. However, C^{sus} with $E_i > 13.00 \mu\text{g}$ on May 1st is 1652 less than that on July 25th. In other words, with the increase of virus transmission time, more and more individuals have higher exposure risks. This trend is consistent with existing cognition and in line with the reality,

TABLE II
RESULTS OF DIFFERENT MODELS

Level	Model	MAE	RMSE	MAPE
Macroscopic	SIR Model	11589.07	13638.09	16.84%
	Grey Model	10594.7	12768.68	15.39%
	Xiao's Model	38965.99	39468.87	58.46%
Microscopic	Hernández-Orallo's Model	34989.95	35537.92	52.41%
	Our Model	8140	9598.17	12.01%

NOTE. MAE, mean absolute error; RMSE, root mean square error; MAPE, mean absolute percentage error.

which demonstrates the validity of the proposed model.

2) *Prediction of New Cases*: To predict the daily new cases C^{new} in the testing set, parameters α^* and β^* should be estimated first. According to the previous results of personal exposure risk, the variation range of α is between 0.50 μg and 100.00 μg . Besides, β is set to change from 1.00×10^{-5} to 2.00×10^{-3} through numerous tests. After expanding the predicted results with the proportion ρ (4.07×10^{-5} in Section III-A), MAE varies with α and β can be seen from Fig. 9. Herein, we find $\alpha^* = 7.00 \mu\text{g}$ and $\beta^* = 6.20 \times 10^{-4}$ (corresponding MAE is 6080.89) based on the training set according to (12). Note that the dark blue curve in Fig. 9(a) indicates that β increases with the growth of α as expected.

Then, with parameters α^* and β^* , the prediction C^{new} in the testing set can be calculated and expanded with the proportion ρ . As a result, our model achieves a good prediction effect under different evaluation indexes: mean absolute error (MAE) is 8140.00, root mean square error (RMSE) is 9598.17, and mean absolute percentage error (MAPE) is 12.01%.

For some machine learning models, the number of training samples with only 60 days is small, which cannot reflect the optimal performance of the model [61]. To verify whether 60 days are enough for training our model, we randomly select several days (e.g., 1 day, 2 days, ..., 59 days) from 60 days as new training sets, and analyze the changes of prediction effects on the same testing set (i.e., the rest 19 days). We estimate α^* and β^* in the same way as before for different new training sets, and the evaluation results on the testing set are shown in Fig. 10. The values of MAE, RMSE, and MAPE decrease with the increase of the number of days in training sets, and each of the indexes converges to a stable value within 60 days. Hence, 60 days are sufficient to be the training set, based on which the parameters obtained are reliable, and the corresponding prediction results can represent the best effect of our model.

IV. COMPARISON WITH OTHER MODELS

In this section, following the study region and period in Section III, we predict transmission trends via our model and several existing models from macroscopic or microscopic levels and then compare their performances.

A. Microscopic-Level Models

Since the proposed model focuses on the general place and determines the personal exposure risk at the microscopic level,

comparison methods should have the same central points, thus Xiao's model [19] and Hernández-Orallo's model [40] are adopted. Besides, as there is no direct prediction of daily new cases in these two comparison models, we follow Section II-D to make the forecast after determining the personal exposure risk by these two models, respectively.

The input data fields required by the two models are the same as our proposed model, and they have the same training and testing sets as Section III. For the contact approach and cough approach of Xiao's model, the cut-off distance of the exposure is 1.00 m and 2.50 m, respectively, as the same settings in [19]. In Hernández-Orallo's model, the contact cut-off distance is also 1.00 m, and the parameter for adjusting the model is set at the same value 1/30 as in [40]. The simulation place is an indoor room, as shown in Fig. 7, which has low air renewal, high temperatures, and low solar radiation, so the value of the quality of the medium in Hernández-Orallo's model is $7/9 = (1+1/3+1)/3$. Based on the training set, we follow (12) to find the α^* and β^* for each model: $\alpha^* = 10.00$ s and $\beta^* = 1.19 \times 10^{-4}$ in Xiao's Model (corresponding MAE is 7205.67); $\alpha^* = 0.60$ MEMs and $\beta^* = 1.41 \times 10^{-4}$ in Hernández-Orallo's Model (corresponding MAE is 7172.78). Hence, based on (9-11), the prediction results of the testing set by the two microscopic-level models are reported in Fig. 11 and Table II.

B. Macroscopic-Level Models

The typical SIR model [21], [62] and grey model [61], [63] are adopted as comparison macroscopic-level models. The SIR model is a traditional infectious disease model, and it works with the assumption that the population N^{PL} in the study region is uniform and homogeneously mixed. In the SIR model, the population is divided into three classes, namely, N^{sus} : susceptible, C^{inf} : infected, and N^{rem} : removed (by recovery and death) [62]. The time-varying number of cases in each class is governed by the infectivity rate σ^i and the removal rate σ^r [64]. Thus, when we know the initial N^{sus} , C^{inf} , N^{rem} , and the constant values of σ^i and σ^r , we can predict the number of infectors in the future, based on which the number of daily new cases C^{new} are calculated. The grey model, favored for its "simple model, strong adaptability, and easy parameter changes", has been widely used in the field of infectious diseases [61]. Unlike the neural network methods that need a substantial number of datasets for training system parameters, the grey model can predict the trend of the COVID-19 well with limited information. Specifically, only the historical daily new infected cases are needed to forecast the C^{new} in the following days.

In Section III, only 60 days (selected from May 1st to July 6th) are used as the training set due to the lack of data. Fortunately, data required by the SIR model and grey model in 67 days (from May 1st to July 6th) are available from the public datasets. Thus, the time-series data from May 1st to July 6th at the national level of the U.S. are adopted as the training set here. The testing set is the same as Section III, i.e., 19 days from July 7th to July 25th. Besides, the parameters of the SIR model and grey model are set based on the national

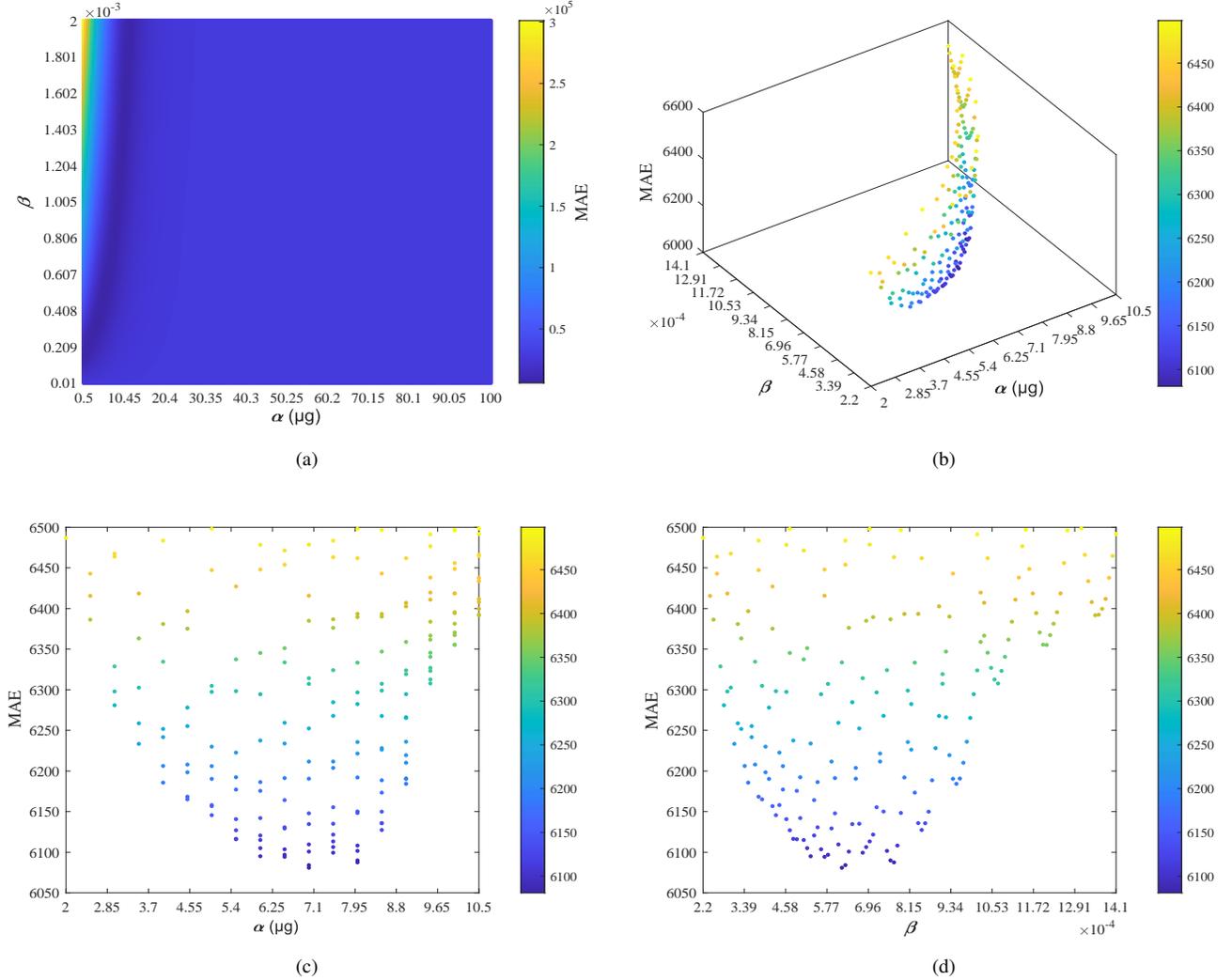


Fig. 9. MAE varies with α and β . (a) The side view of $\alpha - \beta$ when $\alpha \in [0.50, 100.00] \mu\text{g}$ and $\beta \in [1.00 \times 10^{-5}, 2.00 \times 10^{-3}]$. (b-d) The front view and 2 side views when MAE are less than 6500.

dataset. In SIR model, the population size of the U.S. is $N^{\text{PL}} = 327167434$, which comes from the U.S. Department of Transportation's Bureau of Transportation Statistics Trips by Distance – National data product [58], [59]. The cumulative number of infected cases C^{inf} is obtained from the U.S. Centers for Disease Control and Prevention [60]. The number of removed cases (by recovery and death) N^{rem} is provided by the record COVID-19 DATA HUB¹ [65]. The number of susceptible individuals is getting as $N^{\text{sus}} = N^{\text{PL}} - C^{\text{inf}} - N^{\text{rem}}$. After fitting the training set with the ordinary least squares, we get parameters that minimize the sum of squares of errors: $\sigma^i = 0.024$ and $\sigma^r = 0.007$. Besides, the basic reproduction number $R_0 = \frac{\sigma^i}{\sigma^r} = 3.43$ is reasonable with existing research [66], and it is in line with the situation when no epidemic prevention and control policies were implemented in the U.S.. For the grey model, the cumulative number of infected cases

C^{inf} is adopted from the U.S. Centers for Disease Control and Prevention [60], and there are no other parameters to be set. Thus, the prediction results of the testing set by the two macroscopic-level models are reported in Fig. 11 and Table II.

C. Comparison of Results

The time series numbers of daily new cases C^{new} in the testing set estimated by the proposed model, macroscopic models, and obtained via the real-world dataset are shown in Fig. 11 and Table II. Based on these, the proposed model achieves the best prediction performance when compared with the four existing models.

The SIR model and grey model at the macroscopic level cannot determine the personal exposure risk and the number of high-risk exposed people, but these can be estimated by microscopic-level models such as ours. Meanwhile, even though time series data are used for training the SIR model and grey model, the number of samples in the training set may still be too small to guarantee the best performance. However,

¹See websites <https://covid19datahub.io/> for more details about COVID-19 DATA HUB.

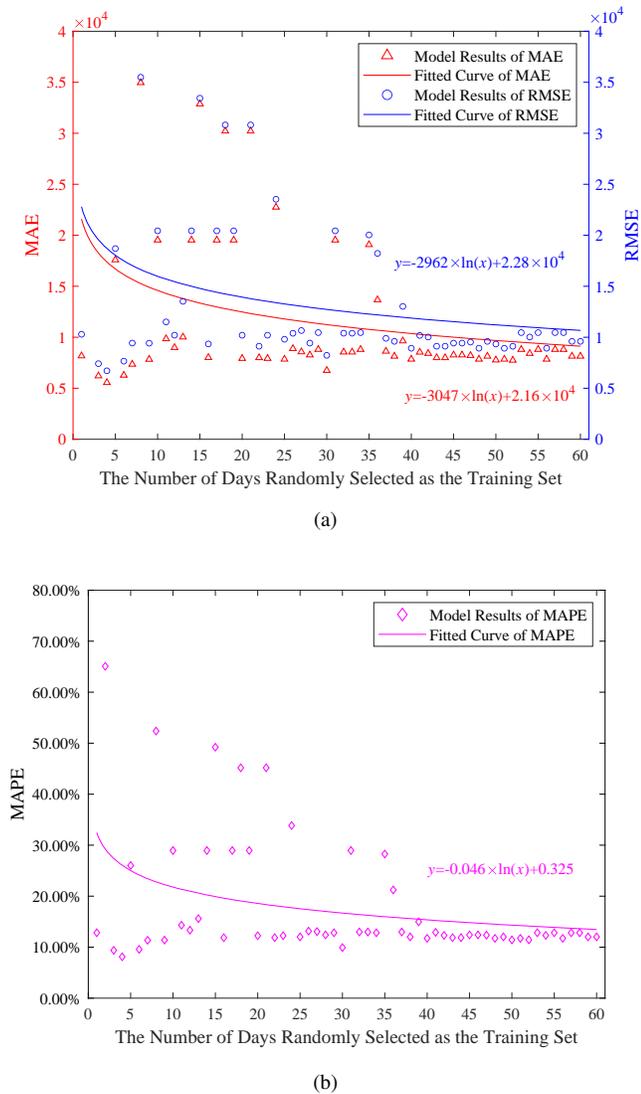


Fig. 10. (a) MAE, RMSE, and (b) MAPE change with the number of days randomly selected as the training set.

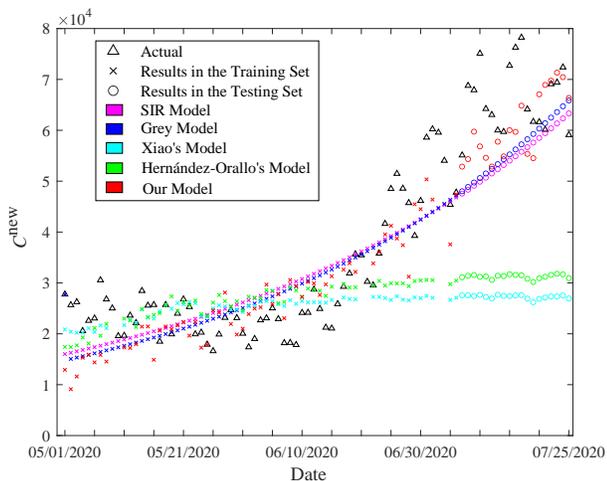


Fig. 11. The actual values and different model results of C^{new} change with the date.

based on the analyses of Fig. 10, the number of samples, in this case, is enough to show the good performance of our proposed model, thus our model is more suitable for small sample data. We admit that these two models are classical and traditional, and now there are many improved ones based on them, which may have better prediction performance than our model.

According to the above descriptions, Xiao's model and Hernández-Orallo's model at the microscopic level differ from our model in determining personal exposure risk. In Xiao's model, there are two ways of virus transmission (i.e., physical contact and cough), and the exposure risk in the influence area is set to a constant value that cannot change with the physical distances. On the other hand, Hernández-Orallo's model only considers the contact infection but adds the dynamic changes of physical distance between people into the study. As shown in Fig. 11 and Table II, Hernández-Orallo's model adopts fewer spreading ways but better prediction effects than Xiao's. These indicate that choosing an appropriate model is more important than considering more transmission routes. Based on this finding, although only the cough infection is considered in our model, it concerns both the time-varying individuals' physical distances and the cough droplet dispersion processes, so as to obtain the best results.

V. CONCLUSION

In this paper, we build a microscopic level exposure-risk-based model to predict the transmission trends of RIDs in a general public place. Specifically, to determine the personal exposure risk, our model couples the motions of individuals and the time-varying cough droplet dispersion process. Then, the number of new cases is predicted based on the assumption that it satisfies a linear function with the number of high-risk exposed people, and the prediction can be tested quantitatively when compared with actual values. Based on COVID-19 data collected in the U.S., we construct real-world simulations and compare the prediction effect of our model with those of several existing models. The results show that our proposed model achieves superior performance than comparison models and possesses significant advantages in a small sample. Moreover, there is an interesting finding: when estimating the personal exposure risk, an appropriate model that describes the dispersion process can make more contributions than adding more transmission routes.

In the application part, a general no-obstacle indoor space is formulated for the reproduction of microscopic epidemic transmission, and by this means, the simulation results of our model can be compared with the macroscopic epidemic transmission data. Indeed, there is a hypothesis in the simulation that the scene geometric characteristics, the activity pattern and many other heterogeneous settings in these different indoor spaces are ignored, and the overall results are obtained by simply expanding the simulation results of the general no-obstacle spaces. It should be noted that our model is able to simulate scenes with different obstacles, activities, and other settings, only if we have the specific scene setting data. Moreover, the potential impact of the hypothesis is minimized since the involved models here are all applied in the same scene.

Nevertheless, there are several limitations of this model. *First*, in the actual scenario, some individual factors of susceptible populations (e.g., age, vaccination) will affect the exposure risk [9], [67], which can be considered in future research. *Second*, to forecast the number of new cases, the assumption that $F(C^{\text{risk}})$ is a linear equation has been illustrated, and other functional forms (e.g., exponential function, power function) can be further studied.

ACKNOWLEDGMENT

The authors would like to thank and acknowledge our partners and data sources in this effort: 1) the daily population traveling out of home and the population size of the U.S. from the U.S. Department of Transportation's Bureau of Transportation Statistics Trips by Distance – National data product; 2) the total number of cases in each day, the cumulative number of infected cases, and the number of daily new infected cases from the U.S. Centers for Disease Control and Prevention; 3) the mean dwell time of individuals from the University of Maryland COVID-19 Impact Analysis Platform; 4) the number of removed cases (by recovery and death) from the dataset named COVID-19 DATA HUB in the URL <https://covid19datahub.io>.

REFERENCES

- [1] F. Araya, "Modeling the spread of covid-19 on construction workers: An agent-based approach," *Safety Science*, vol. 133, p. 105022, 2021.
- [2] R. J. Milne, C. Delcea, and L.-A. Cotfas, "Airplane boarding methods that reduce risk from covid-19," *Safety Science*, vol. 134, p. 105061, 2021.
- [3] A. Kuzdeuov, D. Baimukashev, A. Karabay, B. Ibragimov, A. Mirzakhetov, M. Nurpeiissov, M. Lewis, and H. A. Varol, "A network-based stochastic epidemic simulator: Controlling covid-19 with region-specific policies," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2743–2754, 2020.
- [4] S. Jin, Y. He, K. Yang, Q. Gan, W. Huang, X. Wang, C. Meng, and H. Wang, "The resumption of sports medicine during the covid-19 post-epidemic period: Experiences from wuhan, people's republic of china," *JBJS*, vol. 103, no. 1, pp. 10–14, 2021.
- [5] M. Moslemifard *et al.*, "Hospital diet for covid-19, an acute respiratory infectious disease: An evidence-based protocol of a clinical trial," *Caspian Journal of Internal Medicine*, vol. 11, no. Suppl 1, p. 466, 2020.
- [6] G. Huang and F. Guo, "Loss of life expectancy due to respiratory infectious diseases: Findings from the global burden of disease study in 195 countries and territories 1990–2017," *Journal of Population Research*, May 2021.
- [7] C.-C. Zhu and J. Zhu, "Dynamic analysis of a delayed covid-19 epidemic with home quarantine in temporal-spatial heterogeneous via global exponential attractor method," *Chaos, Solitons & Fractals*, vol. 143, p. 110546, 2021.
- [8] A. Verma, P. Bhattacharya, M. Zuhair, S. Tanwar, and N. Kumar, "Vaccination: Blockchain-based 5g-assisted uav vaccine distribution scheme for future pandemics," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2021.
- [9] S. Daniels, H. Wei, Y. Han, H. Catt, D. W. Denning, I. Hall, M. Regan, A. Verma, C. A. Whitfield, and M. van Tongeren, "Risk factors associated with respiratory infectious disease-related presenteeism: a rapid review," *medRxiv*, 2021.
- [10] D. M. Skowronski, C. Chambers, S. Sabaiduc, G. De Serres, A.-L. Winter, J. A. Dickinson, M. Krajden, J. B. Gubbay, S. J. Drews, C. Martineau *et al.*, "A perfect storm: impact of genomic variation and serial vaccination on low influenza vaccine effectiveness during the 2014–2015 season," *Clinical Infectious Diseases*, vol. 63, no. 1, pp. 21–32, 2016.
- [11] L. C. Ramsay, S. A. Buchan, R. G. Stirling, B. J. Cowling, S. Feng, J. C. Kwong, and B. F. Warshawsky, "The impact of repeated vaccination on influenza vaccine effectiveness: a systematic review and meta-analysis," *BMC medicine*, vol. 17, no. 1, pp. 1–16, 2019.
- [12] E. Mahase, "Covid-19: What new variants are emerging and how are they being investigated?" *BMJ*, vol. 372, 2021.
- [13] S. Lai, N. W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J. R. Floyd, A. Wesolowski, M. Santillana, C. Zhang, X. Du *et al.*, "Effect of non-pharmaceutical interventions to contain covid-19 in china," *nature*, vol. 585, no. 7825, pp. 410–413, 2020.
- [14] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton *et al.*, "Estimating the effects of non-pharmaceutical interventions on covid-19 in europe," *Nature*, vol. 584, no. 7820, pp. 257–261, 2020.
- [15] D. Duives, Y. Chang, M. Sparnaaij, B. Wouda, D. Boschma, Y. Liu, Y. Yuan, W. Daamen, M. C. de Jong, C. Teberg *et al.*, "The multi-dimensional challenges of controlling sars-cov-2 transmission in indoor spaces: Insights from the linkage of a microscopic pedestrian simulation and virus transmission models," *medRxiv*, 2021.
- [16] N. Perra, "Non-pharmaceutical interventions during the covid-19 pandemic: A review," *Physics Reports*, vol. 913, pp. 1–52, 2021.
- [17] J. Ren, Y. Yan, H. Zhao, P. Ma, J. Zabalza, Z. Hussain, S. Luo, Q. Dai, S. Zhao, A. Sheikh *et al.*, "A novel intelligent computational approach to model epidemiological trends and assess the impact of non-pharmacological interventions for covid-19," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3551–3563, 2020.
- [18] E. Hosseini, K. Z. Ghafoor, A. S. Sadiq, M. Guizani, and A. Emrouznejad, "Covid-19 optimizer algorithm, modeling and controlling of coronavirus distribution process," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2765–2775, 2020.
- [19] Y. Xiao, M. Yang, Z. Zhu, H. Yang, L. Zhang, and S. Ghader, "Modeling indoor-level non-pharmaceutical interventions during the covid-19 pandemic: a pedestrian dynamics-based microscopic simulation approach," *Transport Policy*, vol. 109, pp. 12–23, 2021.
- [20] D. J. Earn, "A light introduction to modelling recurrent epidemics," in *Mathematical epidemiology*. Springer, 2008, pp. 3–17.
- [21] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the royal society of london. Series A. Containing papers of a mathematical and physical character*, vol. 115, no. 772, pp. 700–721, 1927.
- [22] H. W. Hethcote, "Qualitative analyses of communicable disease models," *Mathematical biosciences*, vol. 28, no. 3–4, pp. 335–356, 1976.
- [23] W.-m. Liu, H. W. Hethcote, and S. A. Levin, "Dynamical behavior of epidemiological models with nonlinear incidence rates," *Journal of mathematical biology*, vol. 25, no. 4, pp. 359–380, 1987.
- [24] P. C. Cross, P. L. Johnson, J. O. Lloyd-Smith, and W. M. Getz, "Utility of r_0 as a predictor of disease invasion in structured populations," *Journal of the Royal Society Interface*, vol. 4, no. 13, pp. 315–324, 2007.
- [25] A. J. Black and A. J. McKane, "Stochasticity in staged models of epidemics: quantifying the dynamics of whooping cough," *Journal of the royal society interface*, vol. 7, no. 49, pp. 1219–1227, 2010.
- [26] O. Castillo and P. Melin, "Forecasting of covid-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic," *Chaos, Solitons & Fractals*, vol. 140, p. 110242, 2020.
- [27] P. Melin, J. C. Monica, D. Sanchez, and O. Castillo, "Multiple ensemble neural network models with fuzzy response aggregation for predicting covid-19 time series: the case of mexico," in *Healthcare*, vol. 8, no. 2. Multidisciplinary Digital Publishing Institute, 2020, p. 181.
- [28] O. Castillo and P. Melin, "A novel method for a covid-19 classification of countries based on an intelligent fuzzy fractal approach," in *Healthcare*, vol. 9, no. 2. Multidisciplinary Digital Publishing Institute, 2021, p. 196.
- [29] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, 2004.
- [30] M. Z. Bazant and J. W. Bush, "A guideline to limit indoor airborne transmission of covid-19," *Proceedings of the National Academy of Sciences*, vol. 118, no. 17, 2021.
- [31] M. Kriegel, U. Buchholz, P. Gastmeier, P. Bischoff, I. Abdelgawad, and A. Hartmann, "Predicted infection risk for aerosol transmission of sars-cov-2," *MedRxiv*, 2020.
- [32] Y. Arav, Z. Klausner, and E. Fattal, "Theoretical investigation of pre-symptomatic sars-cov-2 person-to-person transmission in households," *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.
- [33] C. X. Gao, Y. Li, J. Wei, S. Cotton, M. Hamilton, L. Wang, and B. J. Cowling, "Multi-route respiratory infection: when a transmission route may dominate," *Science of the Total Environment*, vol. 752, p. 141856, 2021.

- [34] E. Ronchi and R. Lovreglio, "Exposed: An occupant exposure model for confined spaces to retrofit crowd models during a pandemic," *Safety Science*, vol. 130, p. 104834, 2020.
- [35] D. Kim and A. Quaini, "Coupling kinetic theory approaches for pedestrian dynamics and disease contagion in a confined environment," *Mathematical Models and Methods in Applied Sciences*, vol. 30, no. 10, pp. 1893–1915, 2020.
- [36] P. S. Abdul Salam, W. Bock, A. Klar, and S. Tiwari, "Disease contagion models coupled to crowd motion and mesh-free simulation," *Mathematical Models and Methods in Applied Sciences*, pp. 1–19, 2021.
- [37] Z. Fang, Z. Huang, X. Li, J. Zhang, W. Lv, L. Zhuang, X. Xu, and N. Huang, "How many infections of covid-19 there will be in the 'diamond princess'-predicted by a virus transmission model based on the simulation of crowd flow," *arXiv preprint arXiv:2002.10616*, 2020.
- [38] D. R. Parisi, G. A. Patterson, L. Pagni, A. Osimani, T. Bacigalupo, J. Godfrid, F. M. Bergagna, M. R. Brizi, P. Momesso, F. L. Gomez *et al.*, "Social distance characterization by means of pedestrian simulation," *arXiv preprint arXiv:2009.04019*, 2020.
- [39] V. Romero, W. D. Stone, and J. D. Ford, "Covid-19 indoor exposure levels: An analysis of foot traffic scenarios within an academic building," *Transportation Research Interdisciplinary Perspectives*, vol. 7, p. 100185, 2020.
- [40] E. Hernández-Orallo and A. Armero-Martínez, "How human mobility models can help to deal with covid-19," *Electronics*, vol. 10, no. 1, p. 33, 2021.
- [41] A. Bouchnita and A. Jebrane, "A multi-scale model quantifies the impact of limited movement of the population and mandatory wearing of face masks in containing the covid-19 epidemic in morocco," *Mathematical Modelling of Natural Phenomena*, vol. 15, p. 31, 2020.
- [42] A. Bouchnita and A. Jebrane, "A hybrid multi-scale model of covid-19 transmission dynamics to assess the potential of non-pharmaceutical interventions," *Chaos, Solitons & Fractals*, vol. 138, p. 109941, 2020.
- [43] M. Chraïbi, U. Kemloh, A. Schadschneider, and A. Seyfried, "Force-based models of pedestrian dynamics," *Networks & Heterogeneous Media*, vol. 6, no. 3, p. 425, 2011.
- [44] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical review E*, vol. 51, no. 5, p. 4282, 1995.
- [45] T. Kretz, J. Lohmiller, and P. Sukennik, "Some indications on how to calibrate the social force model of pedestrian dynamics," *Transportation research record*, vol. 2672, no. 20, pp. 228–238, 2018.
- [46] L. Liu, Y. Li, P. V. Nielsen, J. Wei, and R. L. Jensen, "Short-range airborne transmission of expiratory droplets between two people," *Indoor air*, vol. 27, no. 2, pp. 452–462, 2017.
- [47] L. Borro, L. Mazzei, M. Raponi, P. Piscitelli, A. Miani, and A. Secinaro, "The role of air conditioning in the diffusion of sars-cov-2 in indoor environments: A first computational fluid dynamic model, based on investigations performed at the vatican state children's hospital," *Environmental research*, vol. 193, p. 110343, 2021.
- [48] S. Zhu, S. Kato, and J.-H. Yang, "Study on transport characteristics of saliva droplets produced by coughing in a calm indoor environment," *Building and environment*, vol. 41, no. 12, pp. 1691–1702, 2006.
- [49] J. K. Gupta, C.-H. Lin, and Q. Chen, "Characterizing exhaled airflow from breathing and talking," *Indoor air*, vol. 20, no. 1, pp. 31–39, 2010.
- [50] *CFD Simulation of Human Coughs and Sneezes: A Study in Droplet Dispersion, Heat, and Mass Transfer*, ser. ASME International Mechanical Engineering Congress and Exposition, vol. Volume 7: Fluid Flow, Heat Transfer and Thermal Systems, Parts A and B, 11 2010.
- [51] N. Gao and J. Niu, "Transient cfd simulation of the respiration process and inter-person exposure assessment," *Building and environment*, vol. 41, no. 9, pp. 1214–1222, 2006.
- [52] L. Zhang and Y. Li, "Dispersion of coughed droplets in a fully-occupied high-speed rail cabin," *Building and Environment*, vol. 47, pp. 58–66, 2012.
- [53] Y. Zhang, G. Feng, Z. Kang, Y. Bi, and Y. Cai, "Numerical simulation of coughed droplets in conference room," *Procedia Engineering*, vol. 205, pp. 302–308, 2017.
- [54] C. Y. H. Chao, M. P. Wan, L. Morawska, G. R. Johnson, Z. Ristovski, M. Hargreaves, K. Mengersen, S. Corbett, Y. Li, X. Xie *et al.*, "Characterization of expiration air jets and droplet size distributions immediately at the mouth opening," *Journal of aerosol science*, vol. 40, no. 2, pp. 122–133, 2009.
- [55] J. Duguid, "The size and the duration of air-carriage of respiratory droplets and droplet-nuclei," *Epidemiology & Infection*, vol. 44, no. 6, pp. 471–479, 1946.
- [56] R. Bi, "A numerical investigation of human cough jet development and droplet dispersion," Master's thesis, Dept. Mech.& Matls.Eng., University of Western Ontario, London, Canada, 2018.
- [57] Z. Li, H. Wang, X. Zhang, T. Wu, and X. Yang, "Effects of space sizes on the dispersion of cough-generated droplets from a walking person," *Physics of Fluids*, vol. 32, no. 12, p. 121705, 2020.
- [58] Maryland Transportation Institute and Center for Advanced Transportation Technology Laboratory at the University of Maryland, "Trips by distance – national," Accessed on: June 27th, 2021. [Online]. Available: <https://data.bts.gov/Research-and-Statistics/Trips-by-Distance-National/6ced-86in>.
- [59] L. Zhang, A. Darzi, S. Ghader, M. L. Pack, C. Xiong, M. Yang, Q. Sun, A. Kabiri, and S. Hu, "Interactive covid-19 mobility impact and social distancing analysis platform," *Transportation Research Record*, vol. 0, no. 0, p. 03611981211043813, 2020.
- [60] Centers for Disease Control and Preventions (CDC), "Covid-19 case surveillance public data access, summary, and limitations," Accessed on: June 22nd, 2021. [Online]. Available: <https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36>.
- [61] J. Zhang and Z. Jiang, "A new grey quadratic polynomial model and its application in the covid-19 in china," *Scientific Reports*, vol. 11, no. 1, pp. 1–27, 2021.
- [62] B. F. Maier and D. Brockmann, "Effective containment explains sub-exponential growth in recent confirmed covid-19 cases in china," *Science*, vol. 368, no. 6492, pp. 742–746, 2020.
- [63] X. Luo, H. Duan, and K. Xu, "A novel grey model based on traditional richards model and its application in covid-19," *Chaos, Solitons & Fractals*, vol. 142, p. 110480, 2021.
- [64] A. Zamiri, H. S. Yazdi, and S. A. Goli, "Temporal and spatial monitoring and prediction of epidemic outbreaks," *IEEE journal of biomedical and health informatics*, vol. 19, no. 2, pp. 735–744, 2014.
- [65] E. Guidotti and D. Ardia, "Covid-19 data hub," *Journal of Open Source Software*, vol. 5, no. 51, p. 2376, 2020.
- [66] Centers for Disease Control and Preventions (CDC), "Covid-19 pandemic planning scenarios," Accessed on: October 20th, 2021. [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html#print>.
- [67] J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani *et al.*, "Changes in contact patterns shape the dynamics of the covid-19 outbreak in china," *Science*, vol. 368, no. 6498, pp. 1481–1486, 2020.